THE WEIL CONJECTURES

BRIAN OSSERMAN

The Weil conjectures constitute one of the central landmarks of 20th century algebraic geometry: not only was their proof a dramatic triumph, but they served as a driving force behind a striking number of fundamental advances in the field. The conjectures treat a very elementary problem: how to count the number of solutions to systems of polynomial equations over FINITE FIELDS. While one might ultimately be more interested in solutions over, say, the field of rational numbers, the problem is far more tractable over finite fields, and LOCAL-GLOBAL PRINCIPLES such as the BIRCH-SWINNERTON-DYER CONJECTURE establish strong, albeit subtle, relationships between the two cases.

Moreover, there are some basic questions that have non-obvious connections to the Weil conjectures. The most famous of these is the *Ramanujan conjecture*, which treats the coefficients of $\Delta(q)$, one of the most fundamental examples of a MODULAR FORM. We obtain the function $\tau(n)$ from the formula for $\Delta(q)$ as follows:

$$\Delta(q) := q \prod_{n=1}^{\infty} (1-q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n.$$

RAMANUJAN conjectured that $|\tau(p)| \leq 2p^{11/2}$ for any prime number p. This is closely related to a statement on the number of ways of writing p as a sum of 24 squares. Work of Eichler, Shimura, Kuga, Ihara, and Deligne showed that in fact Ramanujan's conjecture is a consequence of the Weil conjectures, so that Deligne's proof of the latter in 1974 also resolved the former.

We begin with a brief historical summary of developments prior to WEIL, followed by a more precise description of the statement of his conjectures, and finally a sketch of the ideas behind their proof.

1. AN AUSPICIOUS PROLOGUE

We begin our story with the seminal work of RIEMANN on the classical ZETA FUNCTION, which we recall is defined by the sum

$$\zeta(s) = \sum_{n} \frac{1}{n^s}. \tag{1}$$

EULER had studied this function for real values of s, but Riemann, in his remarkable 8-page paper of 1859, went much further. He looked at complex values as well, and therefore had at his disposal the considerable resources of complex analysis. In particular, although the above sum for $\zeta(s)$ converges only for complex numbers s that have real part $\Re(s)$ strictly greater than 1, Riemann showed that the function itself can be extended to an analytic function defined on the entire complex plane, except at the point $s = 1$, at which it tends to infinity. He showed moreover that $\zeta(s)$ satisfies a certain functional equation relating $\zeta(s)$ to $\zeta(1-s)$, which introduced an important kind of symmetry around the line $\Re(s) = \frac{1}{2}$. Most famously (or infamously), he conjectured what is now known as the RIEMANN HYPOTHESIS, that aside from easily-analyzed “trivial zeroes” on the negative real axis, every zero of $\zeta(s)$ occurs on the line $\Re(s) = \frac{1}{2}$. Riemann’s motivation for studying $\zeta(s)$ was to analyze the distribution of prime numbers, but it fell to later authors (HADAMARD, DE LA VALLÉE-POUSSIN and Van Koch) to bring this vision to fruition. They used the zeta function to prove the PRIME NUMBER THEOREM, which determined the asymptotic distribution of prime numbers, and also showed that the Riemann hypothesis is equivalent to a particularly strong upper bound for the error term in the prime number theorem.

At first glance, the Riemann hypothesis might appear to be completely special, a one-of-a-kind conjecture. However, it was not long before DEDEKIND generalized the Riemann hypothesis to a whole family of zeta functions, and in doing so opened the door to further generalization. Just as we can think of the complex numbers as being obtained from the real numbers by including a square root of -1, that is, a root of the polynomial $x^2 + 1$, a NUMBER FIELD, the fundamental object of study in algebraic number theory, is obtained from the field $\mathbb{Q}$ of rational numbers by including roots of more general polynomials. For each number field K we have the RING OF INTEGERS $\mathcal{O}_K$, which enjoys many of same properties as the classical integers $\mathbb{Z}$. Starting from this observation, Dedekind defined a more general class of zeta functions which now bear his name, associating to each ring of integers $\mathcal{O}_K$ a zeta function in such a way that the classical $\zeta(s)$ is the zeta function for $\mathbb{Z}$. In contrast to the situation for Riemann’s zeta function, the functional equation for Dedekind zeta functions remained open until 1917, when it was settled by Hecke, who showed at the same time that Dedekind zeta functions could be extended to the complex plane, thereby ensuring that the Riemann hypothesis makes sense for them as well.

With such ideas in the air, it was not long before geometry entered the picture. EMIL ARTIN first introduced zeta functions and the Riemann hypothesis for certain curves over finite fields in his 1923 thesis, noting that the ring of polynomial functions on such a curve shares precisely the properties of rings of integers which Dedekind used to define his zeta functions. Artin quickly observed first that his new zeta functions were strongly analogous to Dedekind zeta functions, and second that they were frequently more tractable: evidence for both is provided by the fact that he was able to check explicitly that the Riemann hypothesis was satisfied for a number of specific curves. The difference between the two situations is encapsulated as follows: while in the number field case one can think of the zeta function as counting primes, in the case of a function field the zeta function may be expressed in terms of the more geometric data of counting points on the given curve. In a 1931 paper F. K. Schmidt generalized Artin’s work, and exploited this geometry to prove a strong form of the functional equation for such zeta functions. Shortly thereafter, in a 1933 paper Hasse was able to prove the Riemann hypothesis in the special case of ELLIPTIC CURVES over finite fields.

2. ZETA FUNCTIONS OF CURVES

We now discuss in more detail the definition and properties of zeta functions associated with curves over finite fields, as well as the theorems of Schmidt and Hasse. Let $\mathbb{F}_q$ denote the finite field having q elements, where $q = p^r$ for some prime number p and some positive integer r. The most classical case is when $q = p$, and $\mathbb{F}_p$ is simply the field of integers modulo p. More generally, we can obtain $\mathbb{F}_q$ by adding roots of polynomials to $\mathbb{F}_p$ just as we do with number fields; in fact, a single root of a single irreducible polynomial of degree r will do.

Artin studied a certain class of curves in plane. Here, "plane" means $\mathbb{F}_q^2$, that is, the set of all pairs (x, y) with x and y in $\mathbb{F}_q$. A curve C is simply the subset of these points where some polynomial $f(x, y)$ with coefficients in $\mathbb{F}_q$ vanishes. Of course, if F is any field containing $\mathbb{F}_q$, then the coefficients are also in F, so it makes sense to talk about $C(F)$, the curve in the larger "plane" F^2 defined by the same equation $f(x, y) = 0$. If F is also a finite field, then $C(F)$ is obviously also finite. The finite fields F containing $\mathbb{F}_q$ turn out to be the fields $\mathbb{F}_{q^m}$ for $m \geq 1$. For each $m \geq 1$ let us define $N_m(C)$ to be the number of points in the curve $C(\mathbb{F}_{q^m})$. The sequence $N_1(C), N_2(C), N_3(C), \ldots$ is what we will wish to study.

Given our plane curve C, we can define the *ring of polynomial functions* $\mathcal{O}_C$ of C. This is simply the ring of polynomial functions on the plane (i.e., in two variables), modulo the equivalence relation that two functions taking the same values on C should be considered the same. Formally, $\mathcal{O}_C$ is simply the quotient ring $\mathbb{F}_q[x, y]/(f(x, y))$. Artin's basic observation was that the definition of the Dedekind zeta function could be applied equally well to the ring $\mathcal{O}_C$, yielding a zeta function $Z_C(t)$ associated with C. However, in our geometric context we have the following equivalent and more elementary formula, which explicitly relates $Z_C(t)$ to the number of points over finite fields:

$$Z_C(t) = \exp\left(\sum_{m=1}^{\infty} N_m(C)\frac{t^m}{m}\right). \tag{2}$$

Schmidt generalized Artin's definition to all curves over finite fields, and gave an elegant description of the zeta function for curves, bearing out Artin's observations in the cases he was able to compute. The nicest form of Schmidt's theorem involves restricting to curves satisfying two additional conditions. The first condition is that rather than considering the curve C in the plane, we will want to "compactify" it by considering instead a *projective* curve; we can think of this as adding some "points at infinity", thus increasing $N_m(C)$ slightly. Second, we will want to impose a technical condition of *smoothness* on C, which is analogous to asking that C be a MANIFOLD.

In order to state Schmidt's result, recall that there is a notion of the *genus* g of a smooth projective curve C, which can be defined to be the dimension of the space of DIFFERENTIALS on C, or if C is a complex curve, as the "number of holes" in the space obtained from the analytic topology on C. By extending certain classical results in algebraic geometry to more general fields, Schmidt proved that for a smooth projective curve C over $\mathbb{F}_q$ of genus g, we have

$$Z_C(t) = \frac{P(t)}{(1-t)(1-qt)}, \tag{3}$$

where $P(t)$ is a polynomial with integer coefficients, of degree $2g$. Furthermore, he proved a functional equation in terms of the substitution $t \mapsto 1/qt$. If we set $t = q^{-s}$, this gives a functional equation for the substitution $s \mapsto 1 - s$, as in Riemann's original work. The Riemann hypothesis for C is then the statement that the roots of $Z_C(q^{-s})$ all have $\Re(s) = \frac{1}{2}$, or equivalently, the roots of $P(t)$ all have norm equal to $q^{-1/2}$. It is an elementary observation that this is equivalent to the assertion that $|N_m(C) - q^m + 1| \leq 2g\sqrt{q^m}$, for all $m \geq 1$.

The next step in exploiting the geometric nature of zeta functions of curves is the observation that if F is a finite field containing $\mathbb{F}_{q^m}$, then the points with coordinates in $\mathbb{F}_{q^m}$ are the fixed points of a function called the *Frobenius map*, which is the map Φ_{q^m} that sends a point $(x, y) \in F^2$ to the point (x^{q^m}, y^{q^m}). It is a simple extension of FERMAT'S LITTLE THEOREM that if $t \in \mathbb{F}_{q^m}$, then $t^{q^m} = t$. Moreover, the converse holds: if F is a field containing $\mathbb{F}_{q^m}$, and $t \in F$ satisfies $t^{q^m} = t$, then $t \in \mathbb{F}_{q^m}$. This follows because in any field, and in particular in F, the polynomial $t^{q^m} - t$ can have at most q^m roots, which must then be precisely the elements of $\mathbb{F}_{q^m}$. It immediately follows that a point $(x, y) \in F^2$ is a fixed point of Φ_{q^m} if and only if $(x, y) \in \mathbb{F}_{q^m}^2$. Moreover, it is elementary that $(s+t)^{q^m} = s^{q^m} + t^{q^m}$, if s, t are in any field containing $\mathbb{F}_p$. Because the coefficients of $f(x, y)$ are in $\mathbb{F}_{q^m}$, it follows that if $f(x, y) = 0$, then

$$f(\Phi_{q^m}(x, y)) = f(x^{q^m}, y^{q^m}) = (f(x, y))^{q^m} = 0,$$

so we see that Φ_{q^m} gives a map from C to itself. Thus, one might hope to study $C(\mathbb{F}_{q^m})$ by analyzing more generally what one can say about the fixed points of maps from C to itself. Hasse successfully applied this point of view to prove the Riemann hypothesis in the case $g = 1$, which is to say the case of elliptic curves. Moreover, we will see that this perspective is woven throughout the fabric of the rest of our story, not only inspiring Weil to make his conjectures, but also suggesting the techniques that ultimately led to their proof.

3. Enter Weil

In 1940 and 1941, André Weil gave two proofs of the Riemann hypothesis for curves over finite fields. Or, to be more accurate, he described two proofs: they both relied on fundamental facts in algebraic geometry which had been proved by analytic methods for varieties over the complex numbers, but which remained beyond the reach of the existing foundations in the case of arbitrary base fields. It was largely in order to address this deficiency that Weil wrote his *Foundations of Algebraic Geometry*, which appeared in 1948 and allowed both of his earlier proofs to be made rigorous.

Weil's *Foundations* constituted a watershed in algebraic geometry, as it introduced for the first time a notion of an *abstract* algebraic variety. Previously, a variety was always a global object, in that it was defined by a single collection of polynomial equations, in either affine or projective space. Weil realized that it would be helpful to have a corresponding locally defined concept, so he introduced abstract algebraic varieties, which are obtained by gluing together affine algebraic varieties in much the same way as manifolds in topology are obtained by gluing together open subsets of affine space. The notion of an abstract variety played a fundamental role in formalizing Weil's proofs, and was also an important precursor to Grothendieck's immensely successful theory of schemes (which are discussed in ARITHMETIC GEOMETRY section 3.3).

The following year, in a remarkable paper in the Bulletin of the AMS, Weil went further, studying zeta functions $Z_V(t)$ associated with higher-dimensional varieties V over finite fields, and taking as his definition the formula (2). While the situation is more complicated in this context, the behavior conjectured by Weil was nonetheless strikingly similar, an utterly natural extension of the case of curves:

(i) $Z_V(t)$ is a rational function of t;

(ii) more explicitly, if $n = \dim V$, we can write

$$Z_V(t) = \frac{P_1(t)P_3(t)\cdots P_{2n-1}(t)}{P_0(t)P_2(t)\cdots P_{2n}(t)},$$

where each root of each $P_i(t)$ is a complex number of norm $q^{-i/2}$;

(iii) the roots of $P_i(t)$ are interchanged with the roots of $P_{2n-i}(t)$ under the substitution $t \mapsto 1/q^n t$;

(iv) if V is the reduction modulo p of a variety $\tilde{V}$ defined over a subfield of $\mathbb{C}$, then $b_i := \deg P_i(t)$ is the ith topological BETTI NUMBER of $\tilde{V}$ using the usual topology.

The last part of (ii) is known as the Riemann hypothesis, while (iii) constitutes a functional equation for the substitution $t \mapsto 1/q^n t$. Note that if we return to Schmidt's theorem (3) in the case of curves, the degrees $1, 2g, 1$ of $1-t, P(t), 1-qt$ are also precisely the Betti numbers of a complex curve of genus g.

4. The proof

Weil's conjectures were inspired by a very intuitive topological picture, derived from considering $V(\mathbb{F}_{q^m})$ as the set of fixed points of Φ_{q^m}. Forgetting for the moment that Φ_{q^m} only makes sense over finite fields, if we imagine that V were defined over the complex numbers, then by using the complex topology we could study the fixed points of Φ_{q^m} by the Lefschetz fixed-point theorem, obtaining a formula in terms of the action of Φ_{q^m} on the COHOMOLOGY GROUPS. Indeed, we could conclude almost immediately the factorization in (ii) (and in particular the rationality asserted in (i)), with each factor $P_i(t)$ corresponding to the action of Frobenius on the ith cohomology group, and we would also have $\deg P_i(t)$ given by the ith Betti number of V. Moreover, the functional equation would follow from Poincaré duality.

It was not long before it became clear that such cohomological arguments might become more than just motivation: there could be a cohomology theory for algebraic varieties over finite fields which would mimic the properties of the classical topological theory, and would allow one to prove the Weil conjectures. Such a cohomology theory is now known as a *Weil cohomology*. Serre was the first to seriously attempt to develop such a theory, but with only limited success. In 1960, Dwork provided a brief detour by using p-adic analysis (see LOCAL-GLOBAL PRINCIPLES) to prove the rationality and functional equation. Shortly thereafter, building on comments of Serre and in collaboration with M. Artin, Grothendieck proposed and developed a candidate for a Weil cohomology, the *étale cohomology*. Indeed, he noted that one could in fact extend the list of desired properties of a Weil cohomology in such a way that the Weil conjectures would follow almost immediately. These properties were known but extremely difficult in the classical case, and included the "hard Lefschetz theorem." In a burst of optimism, Grothendieck referred to them as the "standard conjectures," and envisioned that the Weil conjectures would ultimately be proved through them.

However, the final chapter of the story did not go entirely according to Grothendieck's plan. His student Deligne set about working on the problem, and was ultimately able to complete an exceedingly subtle and intricate proof using induction on the dimension of the variety. The étale cohomology played an absolutely fundamental role in Deligne's proof, but he also introduced other ideas into the picture, most notably a classical geometric construction of Lefschetz, as well as some work of Rankin on the Ramanujan conjecture. In the end, he was in fact able to conclude the hard Lefschetz theorem from his work, but the rest of the standard conjectures remain unsolved to this day.

Acknowledgements. I would like to thank Kiran Kedlaya, Nicholas Katz, and Jean-Pierre Serre for their helpful correspondence.

Further reading

Dieudonné, J. 1975. The Weil conjectures. *The Mathematical Intelligencer*, **10**, pp. 7–21.

Katz, N. 1976. An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields. In *Mathematical developments arising from Hilbert problems*, pp. 275-305. Amer. Math. Soc., Providence, RI.

Weil, A. 1949. Numbers of solutions of equations in finite fields, *Bulletin of the AMS*, **55**, pp. 497-508.